

AperTO - Archivio Istituzionale Open Access dell'Università di Torino

Population genetic analyses provide insights on the introduction pathway and spread patterns of the North American forest pathogen *Heterobasidion irregulare* in Italy

This is the author's manuscript

Original Citation:

Availability:

This version is available <http://hdl.handle.net/2318/141100> since 2015-11-24T15:04:59Z

Published version:

DOI:10.1111/mec.12452

Terms of use:

Open Access

Anyone can freely access the full text of works made available as "Open Access". Works made available under a Creative Commons license can be used according to the terms and conditions of said license. Use of all other works requires consent of the right holder (author or publisher) if not exempted from copyright protection by the applicable law.

(Article begins on next page)



UNIVERSITÀ DEGLI STUDI DI TORINO

This is an author version of the contribution:

Questa è la versione dell'autore dell'opera:

*[Garbelotto M., Guglielmo F., Mascheretti S., Croucher P.J.P., Gonthier P., 2013.
Molecular Ecology, 22, pp. 4855–4869, DOI: 10.1111/mec.12452]*

The definitive version is available at:

La versione definitiva è disponibile alla URL:

[<http://onlinelibrary.wiley.com/doi/10.1111/mec.12452/full>]

Population genetics analyses provide insights on the introduction pathway and spread patterns of the North American forest pathogen *Heterobasidion irregulare* in Italy

M. GARBELOTTO, * F. GUGLIELMO, † S. MASCHERETTI, * P. J. P. CROUCHER* and P. GONTHIER†

* University of California at Berkeley, Department of Environmental Science, Policy and Management, 151 Hilgard Hall, Berkeley, CA 94720, USA, † University of Torino, Department of Agricultural, Forest and Food Sciences, Via L. da Vinci 44, I-10095 Grugliasco, Italy

Keywords: autocorrelation analysis, biological invasion, fungal dispersal, genetic diversity, military transport, *Pinus*

Corresponding author: Paolo Gonthier, University of Torino, Department of Agriculture, Forest and Food Sciences, Via L. da Vinci 44, I-10095 Grugliasco, Italy; tel.: +390116708697, fax: +390112368697, e-mail: paolo.gonthier@unito.it

Running title: Invasion genetics of an exotic fungus

Abstract

A population genetic approach is used to identify the most likely introduction site and introduction pathway for the North American forest pathogen *Heterobasidion irregulare* using 101 isolates from six sites in Italy, and 34 isolates from five sites in North America. Diversity indices based on sequences from ten loci indicate the highest diversity in Italy is found in Castelfusano/Castelporziano, and that diversity progressively decreases with increasing distance from that site. AMOVA, Bayesian clustering, and Principal Coordinates analyses based on 12 SSR loci indicate high levels of gene flow among sites, high frequency of admixing, and fail to identify groups of genotypes exclusive to single locations. Cumulatively, these analyses suggest the current infestation is the result of multiple genotypes expanding their range from a single site. Based on two sequenced loci, a single source site in North America could provide enough variability to explain the variability observed in Italy. These results support the notion that *H. irregulare* was introduced originally in Castelporziano: because Castelporziano has been sealed off from the rest of the world for centuries except for a camp set up by the US military in 1944, we conclude the fungus may have been transported in infected wood used by the military. Finally, spatial autocorrelation analyses using SSR data indicate a significant under-dispersion of alleles up to 0.5-10 km, while a significant over-dispersion of alleles was detected at distances over 80 km: these ranges can be used to make predictions on the likely dispersal potential of the invasive pathogen.

Introduction

Emerging animal and plant diseases caused by fungi and fungal-like organisms have steadily increased in the last decades and are predicted to have a significant impact on human and ecosystem health (Fisher *et al.* 2012). In the case of fungal plant pathogens, several epidemics threatening the survival of native plant species have occurred in the last century as a consequence of the introduction of exotic fungi. Examples include, among many other cases, chestnut blight, white pine blister rust, and Dutch elm disease (Maloy 1997; Fisher *et al.* 2012). While multiple introductions from different source populations may reduce the erosion of genetic diversity, invasion is generally associated with a loss of genetic diversity along the expansion gradient as a consequence of the effects of serial bottlenecks (Le Corre & Kremer 1998), and such a pattern has already been demonstrated for plant pathogenic fungi (Vercken *et al.* 2010; Fontaine *et al.* 2013). A range of markers, including rapidly evolving microsatellites, may be used to infer introduction patterns and reconstruct the invasion history of exotic organisms (Estoup & Guillemaud 2010; Fontaine *et al.* 2013). Such a knowledge is important not only for a correct interpretation of the evolutionary processes leading to successful invasions (Dutech *et al.* 2010), but also to improve predictive models and for the implementation of effective and sustainable control strategies against the invasive pathogens (Fontaine *et al.* 2013). The case of the introduction of the fungal forest pathogen *Heterobasidion irregulare* (Underw.) Garbel. & Otrósina (Basidiomycota) from North America into Italy (Gonthier *et al.* 2004) has at least two significant implications with regard to the global movement of infectious agents between continents. First, it provides an unexpected example of a successful introduction of a plant pathogen not known to produce any surviving or resting structures. Second, it is maybe the best-documented introduction of a plant pathogen hypothesized to have occurred through military activities. The widely-accepted belief that *H. irregulare* may have been unwittingly introduced towards the end of World War II by US troops is based upon the fact that *H. irregulare* was first discovered in the hunting grounds of the Presidential Estate of Castelporziano, a forest that has been inaccessible to the general public for centuries (Gonthier *et al.* 2004). The high-walls of the large estate were breached by the 85th Division of the Fifth US Army in June 1944, and a tent camp was subsequently established in proximity of the area where most of the tree mortality is currently observed (D'Amico *et al.* 2007; Gonthier *et al.* 2007), and where discarded wooden crates and latrines were abandoned as the allied troops marched north (Cole 2012). The later discovery that the exotic organism was actually present in pine stands along over 100 km of coastline challenged the solidity of that assumption, and put into

question not only the means, but also the date of the introduction (D'Amico *et al.* 2007; Gonthier *et al.* 2007).

However, two additional observations supported the hypothesis that Castelporziano may have been the site of the original introduction: 1)- disease foci caused by the North American pathogen were about one order of magnitude larger in Castelporziano than anywhere else, and, 2)- the large forest of the Circeo National Park was only infested by *H. irregulare* in its northern portion, suggesting recent colonization from the North, where Castelporziano is located (Gonthier *et al.* 2007; Gonthier *et al.* 2012). However, it should not be overlooked that the validity of the assumption that more severe symptoms are indicative of older infestations is limited by the fact that different ecological conditions among sites may lead to different disease spread rates (Holdenrieder *et al.* 2004).

Given the importance of reconstructing the date, site, and actual pathway of introduction - all factors with clear implications for worldwide policy-making and for adequate prediction of the future rate of spread of this invasive pathogen - further evidence is needed to support or reject the military pathway hypothesis, and to better understand its potential for dispersal within Europe. The North American origin of invasive populations is not in question here: Linzer *et al.* (2008) have solidly reconstructed the phylogeography of North American and Eurasian populations of *Heterobasidion* spp. infecting pines and have shown the Italian population of *H. irregulare*, at that time still referred to as *H. annosum* P intersterility group, to come from eastern North America. Not surprisingly, given the extensive range of *H. irregulare* in the area of origin (Garbelotto & Gonthier 2013), and in spite of the extensive sampling which included populations as far north as Quebec, as far south as Louisiana, and as far west as Missouri, the exact origin of the invasive population could not be ascertained within eastern North America. Nonetheless, Italian sequences of *H. irregulare* did cluster together when a Bayesian analysis of four combined loci was performed, and the DNA sequence of an insertion in the ML5-ML6 mitochondrial region was highly similar only to sequences found in the southeastern USA, suggesting a single provenance from that region (Linzer *et al.* 2008).

The native Eurasian congener *H. annosum* (Fr.) Bref. is present in the entire zone of infestation in Italy but, with the exception of the mesic Circeo forest in the south, its incidence is very low (Gonthier *et al.* 2007). *Heterobasidion irregulare* can also be found in Italian oak woodlands where the native species is virtually absent, but it does not appear to be causing any disease in these stands (Gonthier *et al.* 2012). Jointly, the two species are considered as some of the most damaging pathogens to conifers worldwide (Ottrosina & Garbelotto 2010; Garbelotto & Gonthier 2013). Although the invasive and the native pathogen species display comparable levels of virulence when

inoculated on American and Eurasian pines, the invasive species has a much higher and constant rate of sporulation resulting in transmission and disease rates that are considerably higher than those of the native species (Garbelotto *et al.* 2010). The production of meiospores (basidiospores) is a critical trait for the range expansion of *Heterobasidion* species because mitospores, although abundantly produced *in vitro*, have never been shown to play a role in nature and because secondary tree to tree spread through root contacts is limited to a maximum of a few tens of meters (Stenlid 1985; reviewed in Garbelotto & Gonthier 2013).

One aspect that is intriguing from an evolutionary perspective, and at the same time complicates attempts to reconstruct the history of the introduction and invasion of *H. irregulare*, is its widespread hybridization with *H. annosum*. The two taxa diverged in allopatry approximately 34-41 million years ago (Dalman *et al.* 2010) but their mating systems have remained largely compatible (Stenlid & Karlsson 1991). The introduction of *H. irregulare* in the range once solely occupied by *H. annosum* has fostered massive interspecific hybridization. Approximately 25% of all isolates studied so far have genomes that are admixtures of the two species, but the incidence of admixed genotypes, the direction of allele introgression, and the percentage of the genome that has introgressed from one species into the other, vary depending on site (Gonthier & Garbelotto 2011).

The goals of this study were to investigate the genetic diversity and the genetic structure of pure – i.e. non-admixed – *H. irregulare* populations in different sites using sequence information from ten loci, and allelic polymorphism from 12 SSR loci. Allelic richness and genotypic diversity were estimated and compared among six major infested sites. AMOVA, Bayesian clustering, and Principal Coordinates analyses were performed to provide a picture of the distribution of genetic variation in the zone of infestation, to ascertain the extent of gene flow among sites, and to evaluate the possibility of multiple distinct introductions. Using sequences from two variable loci and 34 individuals from five populations in North America, we determined whether a single North American population could be the source of the entire Italian population. Finally, genetic structure of populations and spatial autocorrelation analyses were used to infer the extent of spatial aggregation of the pathogen and to provide some measurement of its potential ability to spread. Because the pathogen cannot establish itself in purely agricultural or urban settings, we hypothesized that each forest stand represents an “island” of available habitat for the invasive organism. Within this framework, we used a population genetics approach to answer the following questions: 1)- do allelic richness estimates and clustering of alleles confirm a single invasion event starting in Castelporziano? 2)- can we estimate the

likely number of founder individuals and determine whether that number may have come from a single US location?

3)- what are the dispersal abilities of this pathogen as inferred from estimates of population structure and from spatial autocorrelation analyses? The overall rationale behind this study was to test whether population genetics analyses are concordant with a single military-assisted introduction in Castelporziano and therefore support conclusions based on field observations, whilst providing additional parameters to improve our ability to predict and model the progression of an invasive fungus.

Materials and Methods

Isolate selection and generation of molecular markers

A total of 101 haploid isolates of the North American species *H. irregulare* from six distinct pine stands in central Italy was employed for this study (Fig. 1). These stands included: the Castelfusano Pinewood Urban Park (11 isolates) (41°43'22.81"N, 12°18'34.10"E), a site contiguous to the Castelporziano Estate; the Fregene Monumental Pinewood (16 isolates) and the Coccia di Morto Estate (3 isolates) (41°51'25.40"N, 12°11'52.79"E and 41°48'00.50"N, 12°13'40.44"E, respectively), 24 and 17 km from Castelporziano, respectively; the Gallinara pine plantation of Anzio (19 isolates) (41°32'00.89"N, 12°33'32.35"E) and the La Campana pine plantation of Nettuno (11 isolates) (41°30'51.80"N, 12°40'19.95"E), which are 24 and 32 km from Castelporziano; the forest of Sabaudia in the Circeo National Park (41 isolates) (41°19'32.96"N, 13°01'56.21"E), the farthest large infestation (68 km) from Castelporziano. Isolates were a subsample of those used in the study by Gonthier & Garbelotto (2011), all carefully selected to exclude any individual showing significant ($\geq 5\%$) introgression of markers from *H. annosum* in a study on hybridization and introgression employing over 500 loci. Latitude and longitude values were recorded for each isolate at the time of collection.

For molecular analyses, DNA of each isolate was extracted from 20 mg of lyophilized mycelium using the Puregene DNA isolation kit (Gentra). DNA concentration was evaluated using a Nanodrop fluorometer (Thermo Fisher Scientific, Wilmington, DE) and diluted to 10 ng/ μ L before being used as a template for molecular analysis. Sequence analysis was performed on ten loci (Table 1). Primers and PCR conditions used for their amplification were described previously (Kretzer & Bruns 1999; Johannesson & Stenlid 2003; Gonthier *et al.* 2007; Linzer *et al.* 2008; Gonthier & Garbelotto 2011). PCR products were cleaned with the ExoSAP-IT clean-up kit (USB Corp.,

Cleveland, OH, USA). Both forward and reverse strands were sequenced using the above primers and BIGDYE Terminator (version 3.1) chemistry, according to the manufacturer's instructions. Reactions were purified by precipitation, and products were visualized on an ABI3100 (Applied Biosystems) capillary sequencer, following ABI protocols. Sequences were edited and aligned in Sequencher (Gene Codes Corp., Ann Arbor, MI, USA). With the exception of BTUB and GST1 (Table 1), located 956947 bp apart on the same scaffold, all nuclear loci were selected to be on different scaffolds of the *H. irregulare* genome (<http://genome.jgi-psf.org/Hetan1/Hetan1.home.html>) (Olson *et al.* 2012), with an average distance of 1294247 bp, to reduce the probability of linkage among markers (Gonthier & Garbelotto 2011).

The software Sciroko (Kofler *et al.* 2007) with the 'perfect MISA-mode' search setting was used to identify 12 simple sequence repeats (SSRs) in the *H. irregulare* genome. Primers were designed using Primer 3 (http://www.embnet.sk/cgi-bin/primer3_www.cgi) in regions flanking the SSR loci, and annealing temperatures were optimized with a gradient PCR. SSRs were all trimers and hexamers within open reading frames (ORFs), and primers were all designed within conserved exonic regions to minimize mis-priming and ensure repeatable results. The lengths of amplicons always matched the theoretical ones calculated *in silico*, taking into account the length of both amplified flanking regions and a variable number of tri- or hexa-repeats. The SSR loci analyzed, their location in the genome and the primer pairs used for their amplification are reported in Table 2. Universal fluorescent labeling (UFL) was used for the detection of SSR markers (Shimizu *et al.* 2002). Accordingly, PCR reactions were performed in a 10 µL final volume containing: 10 ng of template DNA, 200 µM of dNTPs, 1.5 mM of MgCl₂, 0.2 µM each of tagged forward primer, reverse primer and fluorescently-labeled reporter primer, and 1 U Taq polymerase. The sequence of the reporter primer was 5'-6-carboxyfluorescein (FAM)-GGTGGCGACTCCTGGAG-3' (Shimizu *et al.* 2002). The PCR conditions were as follows: an initial denaturation of 1 min at 95 °C, followed by 35 cycles of denaturation at 94 °C for 30 s, annealing at a temperature ranging from 52°C to 59°C (Table 2) for 30 s, and extension at 72°C for 30 s, and a final extension of 72°C for 5 min. The PCR products were sized on an ABI3100 using Rox 500 as the size marker. For statistical analysis, allele-fragment length was substituted with the number of repeats corresponding to the allele size.

Genetic diversity analyses

Sequence data from 62 *H. irregulare* isolates (Supporting information 1) were employed to calculate indices of diversity for each study site. Because the number of isolates successfully sequenced at multiple loci was relatively small for Fregene and Coccia di Morto, and because Principal Coordinates Analysis (PCoA) (see results) showed: 1)- no association of Coccia di Morto isolates with a particular cluster, 2)- Coccia di Morto isolates were positioned in close proximity to isolates from other sites, and 3)- in two out of three cases isolates from Coccia di Morto were placed close to isolates from Fregene, calculations were performed combining the individuals from the two sites. Fregene and Coccia di Morto are located 6.2 km from one another on the same bank of the Tiber river, while Castelfusano is 17 km to the south-east and on the opposite side of the Tiber.

Calculations were performed for each sequence separately, before being averaged across loci. Only variable loci were used to calculate averages. Two indices were calculated using Arlequin software v.3.5 (Excoffier *et al.* 2005): haplotype diversity (H) and gene diversity (\hat{H}) (Nei & Tajima 1981; Nei 1987).

In order to determine whether the observed diversity in Italy may have been the result of the introduction of genotypes from a single population in North America, the same indices were calculated using the two highly polymorphic loci EFA and GPD using 34 individuals from five North American *H. irregulare* populations and for Castelfusano, the most diverse population in Italy (see results). North American populations were sampled previously (Linzer *et al.* 2008) and were: Lac La Blanche and Cushing (Quebec, Canada), Forêt Larose (Ontario, Canada), Savannah river (SC, USA) and Georgia/Alabama (USA). For further details, including GenBank accessions and coordinates, refer to Linzer *et al.* (2008).

Pearson correlation ($P < 0.05$) and linear regression analyses were performed using PASW Statistics 18.0 (SPSS Inc., Chicago, IL, USA) for each diversity index against the distance (km) from Castelfusano. Due to its proximity to Castelporziano, this site was inferred to be the site among those studied with the oldest infestation of *H. irregulare*.

In order to compare the different sites standardizing sample size, allelic richness (i.e. number of distinct alleles expected in random subsamples of different sizes drawn from the population) and private allelic richness (i.e. number of private alleles expected in the population when random subsamples of different sizes are taken from each of the populations under consideration) were calculated through rarefaction analysis using the ADZE software v. 1.0 (<http://rosenberglab.bioinformatics.med.umich.edu/adze.html>).

The software package Bottleneck v.1.2.02 (Piry *et al.* 1999) was used to determine which of the five Italian populations exhibits an observed gene diversity (H_e) significantly higher than the expected equilibrium gene diversity (H_{eq}), which is computed from the observed number of alleles, under the assumption of a constant-size (equilibrium) population. Gene diversity excess ($H_e > H_{eq}$) is typical of recently bottlenecked populations (Luikart *et al.* 1998). This analysis was performed on polymorphic loci and estimation was based on 10000 replications; significant excess of gene diversity ($P < 0.05$) was assessed through the one-tailed Wilcoxon test under the assumption of loci evolving with the Infinite Allele Model (IAM) and population at mutation–drift equilibrium.

Genetic structure and spatial autocorrelation analyses

SSRs loci were employed to analyze the genetic structure of 98 *H. irregulare* isolates from six sites (Supporting information 2). The Bayesian clustering and assignment software STRUCTURE v. 2.3.3 (Pritchard *et al.* 2000; Falush *et al.* 2007; Hubisz *et al.* 2009) was used to identify the presence of genetically distinguishable subpopulations within the entire sample studied. Because in natural *Heterobasidion* infestations, all genotypes further than a few tens of meters are generated by sexually formed basidiospores (Garbelotto & Gonthier 2013), the genome admixing model for uncorrelated allele frequencies was used. We tested for up to six clusters (K), with five repetitions for each K , to match the total number of study sites. A 100,000 run burn-in and MCMC analyses based on 100,000 interactions were used without assuming prior population information. Selection of K from this output data was carried out based on suggestions in the software documentation, i.e. using the maximum likelihood estimate method suggested by Evanno (Evanno *et al.* 2005). Output files for the optimum estimate of K were merged in CLUMPP (Jakobsson & Rosenberg 2007) and visualized using DISTRUCT (Rosenberg 2004).

A PCoA was also run with R software (R Core Team 2012) with a *labdsv* library (Roberts 2012) to provide more information on the number, composition, and relatedness of distinct clusters. An Euclidean distance matrix was calculated on the original variables space with the *dist()* function and its default options for missing values (Crawley 2013). A Hierarchical Cluster Analysis (HCA) with Euclidean distance and Ward agglomerative method was run on the isolates principal coordinates. The resulting dendrogram, the agglomeration distance (Muñoz-Díaz & Rodrigo 2004), the connectivity and the average proportion of non-overlap (APN) (Brock *et al.* 2008) were examined at each clustering step to find a reasonable number of neighbouring points clusters (i.e. groups of similar isolates). To

confirm the results of HCA, kernel density analysis was performed on principal coordinates (Duong *et al.* 2008). An asymmetric kernel bandwidth was chosen according to the values provided by the smoothed cross validation selector (SCV) (Duong 2007). Kernel density estimates were represented on a raster surface with the indication of significant feature (P -value set at 5%) derived by the curvature-based test (Duong *et al.* 2008). Each HCA cluster was regarded as validated when a significant maximum was observed within the cluster itself.

An analysis of molecular variance (AMOVA; Excoffier *et al.* 1992), as implemented by Arlequin v.3.5, was performed to study the presence of significant genetic structure within and among study sites. Pairwise Φ_{ST} values were calculated for each combination of sites and their significance was calculated after a Bonferroni correction for multiple comparisons.

Finally, Spagedi v.1.3 (Hardy & Vekemans 2002) was used to determine the presence of significant clustering of alleles using spatial autocorrelation. Spatial coordinates for each individual were supplied in UTM format (WGS84). Moran's index of genetic similarity, I (Sokal & Oden 1978), was used to test the correlation between allelic repeats and spatial distance classes. A total of 20000 permutations were used for significance testing of Moran's I .

Results

A total of 33 multilocus genotypes out of 34 isolates successfully sequenced at least at two loci were detected from sequence data (Supporting information 1), and 89 multilocus genotypes out of 98 isolates were detected from SSR data (Supporting information 2). However, the combined analysis of sequence and SSR data indicated all isolates were different multilocus genotypes.

Genetic diversity analyses

Seven of the ten loci sequenced proved to be polymorphic, with a number of alleles per locus ranging between two and six (Table 1). With the exception of Castelfusano, where we detected up to five alleles per sequenced locus, the greatest number of alleles per locus was four in all the other pine stands (Supporting information 3). A list of all alleles per locus including GenBank accession numbers is reported in Supporting information 1.

Both average diversity indices showed highest diversity values for Castelfusano, intermediate values for Anzio, Nettuno and Fregene-Coccia di Morto, and lowest values for the Sabaudia forest at Circeo (Fig. 2). Each diversity

index decreased significantly with increasing distance from Castelfusano (Pearson's r : $P < 0.05$), following the linear regression models reported in Fig. 2. Diversity indices for each polymorphic locus per site are reported in the Supporting information 4.

Rarefaction analysis indicated the highest and lowest allelic richness values were in Castelfusano and in the Sabaudia forest at Circeo, respectively (Fig. 3a). Private allelic richness was also highest in Castelfusano, with values rising with increasing sample size (Fig. 3b).

When indices were calculated using the two loci EFA and GPD for Castelfusano, the most diverse site in Italy, and for five distinct sites in North America, results indicated that at least two of the North American sites contained as much diversity as Castelfusano (Supporting information 4).

Bottleneck analysis showed a significant gene diversity excess ($H_e > H_{eq}$) in Castelfusano ($P = 0.0391$), Anzio ($P = 0.0391$), and Nettuno ($P = 0.0273$), but not in Circeo ($P = 0.5312$) and Fregene-Coccia di Morto ($P = 0.5938$).

Genetic structure and spatial autocorrelation analyses

Ten of the 12 SSR loci analyzed were polymorphic, with the number of alleles per locus ranging between two and seven overall, or between one and six on a site-by-site basis (Tables 2 and Supporting information 3). The complete SSR dataset is available in the Supporting information 2. Because of the high frequency of null alleles, SSR locus 29a was not used for any of the analyses.

When performing the STRUCTURE analysis, the clustering level $K = 3$ yielded the largest delta- K value, followed by $K = 2$ and $K = 4$ (Supporting information 5). At $K = 3$, 2, and 4 individuals assigned to all clusters were present at each study site (Fig. 4a). At all sites there was a prevalence of admixed genotypes, belonging to more than one cluster.

In the PCoA (Fig. 4b), two principal axes were selected explaining 69.9% (principal axis 1) and 13.6% (principal axis 2) of the variance. The HCA dendrogram, the agglomeration distance, the connectivity, and the average proportion of non-overlap at each clustering step showed 2 to 4 clusters of isolates may be present (Supporting information 6). Regardless of the number of clusters, each one included isolates from different populations (Fig. 4b). Kernel density analysis showed 3 distinct significant maxima, corresponding to areas in the principal axis space occupied by 3 different clusters.

AMOVA results (Table 3) indicated most of the genetic variance (93.70%) to be within-site, while a relatively small albeit significant fraction of the genetic variance (6.30%, $P = 0.0429$; Table 3) was detected among sites. When pairwise comparisons were made, and significance threshold was Bonferroni-corrected for multiple comparisons, the only significantly different pair was Castelfusano-Circeo, although all pairwise comparisons between the Circeo and other sites were approaching significance (Table 4).

Spatial autocorrelation analyses indicated a very significant ($P < 0.001$) under-dispersion of alleles between 0 and 500 m, and significant under-dispersion ($P < 0.05$) until 10 km (i.e. individuals within these distance classes were genetically more similar than expected by chance). A very significant ($P < 0.001$) over-dispersion of alleles was detected at distances over 80 km (Fig. 5).

Discussion

A meta analysis of all published data has recently highlighted the steady rise of the introductions of invasive forest pathogens (IFPs) into Europe since the end of World War I (Santini *et al.* 2013). While most IFPs are ascomycetes (kingdom: Fungi) and oomycetes (kingdom: Stramenopila) introduced through infected live plant material, about 10% of all IFPs are basidiomycetous fungi, and infected wood ranks as one of the top three most common introduction pathways (Santini *et al.* 2013). The case of the transport of the basidiomycete *H. irregulare* from North America into Europe provides an excellent opportunity to study in detail the invasion dynamics of a species belonging to a group of microbes that - with the exception of the rusts - has only been studied to a limited extent. While a few other exotic necrotrophic basidiomycetes have been reported (Gonthier 2010), including *Armillaria mellea* (Vahl) P. Kumm. in South Africa (Coetzee *et al.* 2001) and *Phaeolus schweinitzii* in Australia (Simpson & May 2002), this may be the clearest case of long-distance movement, establishment and successful invasion of a wood decay basidiomycete.

As for all biological introductions, a genetic bottleneck is inevitable when IFPs are introduced, but the reproductive biology of the introduced organism, the history of each introduction, and the presence of interfertile native species are all factors that play an essential role in shaping the nature of the invasion itself, and in affecting our ability to correctly reconstruct the invasion dynamics of the exotic species (Garbelotto 2008). Cases are known of multiple introductions of the same pathogen species, but from different source populations (Dutech *et al.* 2012). This scenario

may lead to the coexistence of genotypically and phenotypically distinct populations if sexual recombination is not common, due to the presence of a single mating type in the zone of invasion, to a homothallic (i.e. selfing) reproductive strategy of the invasive species, to environmental conditions or other factors (e.g. viruses) possibly favoring the expansion of clonal lineages (Dobrowolski *et al.* 2003; Engelbrecht *et al.* 2004; Gordon *et al.* 2006; Dutech *et al.* 2010). Where heterothallism (i.e. outcrossing) and multiple mating types are present, the presence of distinct founder populations can allow for recombination events that may accelerate the adaptive process of the invasive species to the new environment (Brasier 2001; Bengtsson *et al.* 2012). At times, as it has been recently shown for the Sudden Oak Death pathogen *Phytophthora ramorum* Werres, De Cock & Man in 't Veld, multiple introductions of the same genotype in distinct locations may lead to the generation of distinct populations, each derived by the independent generation of novel successful genotypes in different areas (Croucher *et al.* 2013). For IFPs that exclusively reproduce asexually in the area of invasion, such as *P. ramorum*, *Ceratocystis platani* (J.M. Walter) Engelbr. & T.C. Harr. or *Seiridium cardinale* (W.W. Wagener) B. Sutton & I.A.S. Gibson (Engelbrecht *et al.* 2004; Della Rocca *et al.* 2011; Croucher *et al.* 2013), the microevolution of invasive population can be simply tracked as the result of genetic divergence of populations due to the accumulation of mutations or somatic recombinations. Conversely, reconstructing the history of the invasion of *H. irregulare* in Italy is rather complex due to the following: 1)- the exact source of the invasive population in the eastern US is not known, thus a precise term of reference is lacking; 2)- populations expand their range exclusively through sexual meiospores, as also strongly supported by the results of the analysis of multilocus genotypes in this study, thus we should expect high levels of recombination which potentially may have masked the signature of original founder individuals; 3)- high levels of hybridization with the congener *H. annosum* may be a source of added allelic diversity that needs to be taken into account in the analyses.

This study circumvents these three issues through an integrated approach based on the following strategies: 1)- selection of individuals without any signs of introgression at over 500 AFLP loci combined with the use of sequence loci that can unambiguously be assigned to *H. irregulare* for estimates of diversity (Linzer *et al.* 2008; Dalman *et al.* 2010; Gonthier & Garbelotto 2011), while using SSR loci within ORF reported often to have alleles that are identical between the two species (Rocchetti 2012) for genetic structure and spatial autocorrelation analyses. Because the reconstruction of the history of the invasive population is strongly based on the extent and distribution of allelic

variability within *H. irregulare* in Italy, it is essential to first exclude any variability originated through allelic introgression from *H. annosum* as that may confound the calculations. Gonthier and Garbelotto (2011) have shown that even in individuals assigned to *H. irregulare* using over 500 AFLP loci, individual genes may still be, thus the use of sequence loci that can be unequivocally assigned to *H. irregulare* is of paramount importance to ensure diversity estimates are correct. On the other hand, the current genetic structure of *H. irregulare* populations is the result of the dispersal and local frequency of both native alleles and alleles introgressed from *H. annosum*, hence the inclusion of both types of alleles may be informative for that purpose; 2)- the use of different approaches to estimate diversity may appear redundant but provides significant confidence to the interpretation of the data. Based on the life cycle of the pathogen and on the results of the analysis of multilocus genotypes, each individual represents a different genotype; this is true for any study such as this one, where adjacent trees possibly infected by the same genotype through vegetative growth (Garbelotto & Gonthier 2013) were never sampled. However, estimates of haplotype diversity are also a function of the number and variability of genetic loci utilized in a study as markers, hence it is a useful metric to include as it allows to determine the power of a specific analytical approach. Haplotype diversity is the measurement more likely to be affected by intergenic recombination processes occurring during sexual reproduction, and thus it is positively correlated with effective population size. Differences in values of haplotype diversity could thus be the result of significant different histories of populations after establishment, rather than being indicators of the nature of founding individuals. Conversely, the gene diversity index is by definition less sensitive to the pure demographic of an invasive species. The generation of a new functional allele is in fact less likely to occur than the generation of a new unique combination of alleles necessary to generate a different genotype. Finally, indices calculated using rarefaction are useful when sample sizes are small as they prevent errors due to unequal evenness of alleles in the sample size (Grünwald *et al.* 2003).

All indices, including richness estimates based on rarefaction, indicate Castelfusano and the Circeo to be the most and least diverse sites, respectively. The two sites are also the only ones for which pairwise values of PHist were significantly different in spite of the fact they are not the two most distant sites, suggesting that they may represent the oldest and youngest infestations in Italy.. PHist values between the Circeo and all other sites approach significance, possibly indicating a recent establishment of *H. irregulare* in the Circeo area and the lack of sufficient time for the population in this site to reach equilibrium with populations elsewhere.

Although inferences on the site of introduction based on genetic diversity may sometimes be erroneous since high levels of genetic diversity can also result from the meeting of different introduction waves (Dutech *et al.* 2010), further lines of evidence, all discussed below, suggest the invasion of *H. irregulare* started in Castelfusano/Castelporziano and ended in the Circeo. In agreement with a recent establishment of the pathogen in the Circeo forest, disease severity is also lowest in this area (Gonthier *et al.* 2007). The fact that only the northern portion of the Circeo forest is colonized by the exotic pathogen (Gonthier *et al.* 2012) also indicates that it may have arrived only recently from the north, and that the time lapsed since its arrival has not been long enough to allow for the colonization of the entire forest stand, 7.3 km in length.

Genetic data supporting the pathogen introduction in Castelfusano/Castelporziano are important as they would provide strong albeit indirect evidence that the military can unwittingly introduce a plant pathogen, maybe by having brought and left infected wood at this site. The potential of long distance spread of plant pathogens through infected wood is a threat discussed in a recent paper (Vannini *et al.* 2012), but lacking concrete evidence in its support.

If we assume a genetic bottleneck in conjunction with the intercontinental movement of founder *H. irregulare* genotypes and lack of massive hybridization with the native *H. annosum* during establishment, the first founding *H. irregulare* population should display diminished allelic diversity (Nei *et al.* 1973; Luikart & Cornuet 1998), but should be characterized by an observed heterozygosity higher than that expected (Luikart *et al.* 1998). Results of the Bottleneck analysis indicated this to be the case for Castelfusano, Anzio, and Nettuno but not for the Fregene-Coccia di Morto or the Circeo sites, also confirming the two latter sites are the result of a colonization process occurring more recently, when allelic diversity and heterozygosity levels had had sufficient time to be balanced.

If infestations are not all started simultaneously, but the pathogen progressively disperses from an initial infestation to other sites, the obvious result is that each new infestation will result in a further reduction of allelic diversity at a rate that will be difficult to be counteracted by the relatively slow generation of new diversity through mutations and recombination events (Nei *et al.* 1973; Hedrick 1999). Our measures of haplotype and genic diversity, as well as of allelic and private allelic richness are all in agreement not with a simultaneous, but with a progressive invasion process starting in Castelporziano and ending at the Circeo. This hypothesis is supported by the significant inverse correlation between all diversity indices and distance from Castelfusano/Castelporziano, as displayed by the linear regression analysis. Regressions calculated assuming other sites as the introduction site were either not significant or

displayed lower R^2 values compared to the ones reported in Fig. 2 (not shown). Incidentally, the severity of symptoms, possibly in direct correlation with age of each infestation, has been described following a similar pattern with most severe symptoms in Castelfusano and least severe symptoms in the Circeo (Gonthier *et al.* 2007).

As expected for a sexually reproducing species, the determination of genetically distinct clusters was not clearcut. This is in contrast with what has been reported for clonally reproducing invasive pathogens, such as *P. ramorum*, where lack of recombination allows for new polymorphisms to be tracked and used to reconstruct the history of the invasion (Croucher *et al.* 2013). Both STRUCTURE and PCoA analyses suggested 3 clusters to be the best option, with 2 and 4 as second and third best, respectively. However, STRUCTURE analysis also revealed a significant level of admixing, confirming abundant sexual reproduction among genotypes and making the precise definition of distinct clusters rather tentative. Nonetheless, number of alleles found in the invasive population can also be used to support a certain number of founder individuals. A maximum of 6 alleles was found per sequenced locus in populations of *H. irregulare* in Italy, while 7 alleles were found for a single SSR locus. These results indicate that 3 or 4 heterokaryotic (n+n) genotypes may have been the source of the infestation, based on sequence and SSR data, respectively. However, while all sequenced alleles are North American in origin (Gonthier & Garbelotto 2011), SSR alleles may either have been introgressed from *H. annosum* or have been generated after the introduction in Italy (Roccatelli 2012), and thus may incorrectly inflate the estimate. The determination of an exact number of clusters or of founder sources is not crucial here, but their distribution is nonetheless very informative. The fact that all clusters are present in all locations, and the lack of clusters exclusively found in one or two sites, are in fact two results best explained by the introduction of genotypes in a single location, followed by their spread to other locations.

Finally, our comparative analysis between Castelfusano and five populations from eastern North America showed that for all indices of diversity employed in this study, two out of five sites contained levels of diversity comparable to those measured in Italy. Such findings would reinforce the notion that the current infestation in Italy is the result of an introduction in a single location from a single source.

Results of the spatial autocorrelation analyses may provide additional valuable information to estimate the spread potential of an invasive organism (Fitzpatrick *et al.* 2012). Caution should be used when interpreting these results because of both spurious associations (Cushman & Langduth 2010) and biases due to sampling schemes (Vekemans & Hardy 2004). Additionally, results may be confounded by the fact that not all populations appear to be at

equilibrium, as indicated in the study of dispersal of the Chestnut blight fungus in North America (Milgroom *et al.* 1996). While results should be interpreted with caution, distances of 500 m should be regarded as easily crossed by the organism. It is interesting to note that a recent study on the patterns of dispersal of the sister taxon *H. annosum* in the same region indicated most spore dispersal occurred within 500 m from a source (Gonthier *et al.* 2012).

Likewise, PHist values confirm the ability of the organism to cross gaps of 10 km when populations are in equilibrium or approach it. Results from the spatial autocorrelation analyses suggest that in regions where available habitats are 500 m-10 km apart (such as the pine stands stretching from Tuscany to northeastern Spain, or the Scots pine plantations of central and northern Europe), the spread rate of the fungus may be higher than the 1.3 km yr⁻¹ estimated by Gonthier *et al.* (2007) in the fragmented habitats of the zone of infestation, which is likely to result in a much faster rate of colonization of the European continent.

Besides suggesting a potential dispersal range for the invasive pathogen, this study highlights the presence of three subtly distinct areas in its zone of infestation: 1)- the zone of initial infestation represented by Castelfusano/Castelporziano and adjoining sites to the south (Anzio and Nettuno), characterized by an excess heterozygosity as a signature of the founding bottleneck effect and lack of structure among sites; 2)- the zone of initial infestation expanded to include Coccia di Morto and Fregene to the north, in which all populations are also without structure, but in which the most recent infestations do not display the characteristic level of excess heterozygosity; and, 3)- the entire zone of infestation including the southernmost Circeo area in which habitat fragmentation and/or (more plausibly) lack of equilibrium due to the recent arrival of the pathogen, currently result in moderate genetic differentiation of the most recent infestation (Circeo) and lack of bottleneck-induced excess heterozygosity. We predict genetic differentiation between the Circeo and the rest of the zone of infestation will decrease in time, as populations reach equilibrium.

We have previously reported that gene introgression from the native congener *H. annosum* into the invasive *H. irregulare* is massive (Gonthier & Garbelotto 2011): the combination of smaller vegetation gaps in central and northern Europe combined with the abundant gene flow documented in this study, may result not only in faster spread of the invasive pathogen but also in a more rapid adaptation to different European habitats, were it to be fostered by the interspecific introgression of useful adaptive genes.

Acknowledgements

We are grateful to Sergio Zerunian for assistance with sampling in the Circeo National Park and to the City of Fregene for assistance with sampling in the local park. We thank Guglielmo Lione for assistance with statistical analyses. We also wish to acknowledge the anonymous reviewers for their valuable and helpful comments on the manuscript. This study is dedicated to the memory of Giovanni Nicolotti who played a critical role in the discovery and early research of the introduction of *Heterobasidion irregulare* into Italy.

Data accessibility

Sample locations and microsatellite data: Supporting information 2 and DRYAD digital repository entry - doi: 10.5061/dryad.m9d40

DNA sequences: GenBank accessions JX982607 – JX982621

Detailed information related to GenBank accession numbers is provided in Table 1 and in Supporting information 1
R script is provided in Supporting information 6

References

- Bengtsson SBK, Vasaitis R, Kirisits T, Solheim H, Stenlid J (2012) Population structure of *Hymenoscyphus pseudoalbidus* and its genetic relationship to *Hymenoscyphus albidus*. *Fungal Ecology*, 5, 147-153.
- Brasier CM (2001) Rapid evolution of introduced plant pathogens via interspecific hybridization. *BioScience*, 51, 123-133.
- Brock G, Pihur V, Datta S, Datta S (2008) clValid: an R package for cluster validation. *Journal of Statistical Software*, 25, 1-22.
- Cole S (2012) History of the 85th 'Custer' Division. Based on a booklet entitled Minturno to the Appennines – 85th Infantry Division. <http://www.custermen.com/History85.htm>. Accessed 29 November 2012.
- Coetzee MPA, Wingfield BD, Harrington TC, Steimel J, Coutinho TA, Wingfield MJ (2001) The root rot fungus *Armillaria mellea* introduced into South Africa by early Dutch settlers. *Molecular Ecology*, 10, 387-396.
- Crawley MJ (2013) *The R Book – Second Edition*. John Wiley & Sons, Ltd, Chichester, UK, pp. 1051.

- Croucher P, Mascheretti S, Garbelotto M (2013) Combining field epidemiological information and genetic data to comprehensively reconstruct the invasion history and the microevolution of the Sudden Oak Death agent *Phytophthora ramorum* (Stramenopila: Oomycetes) in California. *Biological Invasions*, in press.
- Cushman SA, Langduth EL (2010) Spurious correlations and inference in landscape genetics. *Molecular Ecology*, 19, 3592-3602.
- Dalman K, Olson Å, Stenlid J (2010) Evolutionary history of the conifer root rot fungus *Heterobasidion annosum* sensu lato. *Molecular Ecology*, 19, 4979-4993.
- D'Amico L, Motta E, Annesi T, Sciré M, Luchi N, Hantula J, Korhonen K, Capretti P (2007) The North American P group of *Heterobasidion annosum* s.l. is widely distributed in *Pinus pinea* forests of the western coast of central Italy. *Forest Pathology*, 37, 303-320.
- Della Rocca G, Eyre CA, Danti R, Garbelotto M (2011) Sequence and simple-sequence repeat analyses of the fungal pathogen *Seiridium cardinale* indicate California is the most likely source of the Cypress canker epidemic for the Mediterranean region. *Phytopathology*, 101, 1408-1417.
- Dobrowolski MP, Tommerup IC, Shearer BL, O'Brien PA (2003) Three clonal lineages of *Phytophthora cinnamomi* in Australia revealed by microsatellites. *Phytopathology*, 93, 695-704.
- Duong T (2007) Ks: kernel density estimation and kernel discriminant analysis for multivariate data in R. *Journal of Statistical Software*, 21, 1-16.
- Duong T, Cowling A, Koch I, Wand MP (2008) Feature significance for multivariate kernel density estimation. *Computational Statistics and Data Analysis*, 52, 4225-4242.
- Dutech C, Fabreguettes O, Capdevielle X, Robin C (2010) Multiple introductions of divergent genetic lineages in an invasive fungal pathogen, *Cryphonectria parasitica*, in France. *Heredity*, 105, 220-228.
- Dutech C, Barrès B, Bridier J, Robin C, Milgroom MG, Ravigné V (2012) The chestnut blight fungus world tour: successive introduction events from diverse origins in an invasive plant fungal pathogen. *Molecular Ecology*, 21, 3931-3946.
- Engelbrecht CJB, Harrington TC, Steimel J, Capretti P (2004) Genetic variation in eastern North American and putatively introduced populations of *Ceratocystis fimbriata* f. *platani*. *Molecular Ecology*, 13, 2995-3005.

- Estoup A, Guillemaud T (2010) Reconstructing routes of invasion using genetic data: why, how and so what? *Molecular Ecology*, 19, 4113-4130.
- Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, 14, 2611-2620.
- Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, 131, 479-491.
- Excoffier L, Laval G, Schneider S (2005) Arlequin ver. 3.0: an integrated software package for population genetics data analysis. *Evolutionary bioinformatics*, 1, 47-50.
- Falush D, Stephens M, Pritchard JK (2007) Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Molecular Ecology Notes*, 7, 574-578.
- Fisher MC, Henk DA, Briggs C, Brownstein JS, Madoff L, McCraw SL, Gurr S (2012) Emerging fungal threats to animal, plant and ecosystem health. *Nature*, 484, 186-194.
- Fitzpatrick BM, Fordyce JA, Niemiller ML, Reynolds RG (2012) What can DNA tell us about biological invasions? *Biological Invasions*, 14, 245-253.
- Fontaine MC, Austerlitz F, Giraud T, Labbé F, Papura D, Richard-Cervera S, Delmotte F (2013) Genetic signature of a range expansion and leap-frog event after the recent invasion of Europe by the grapevine downy mildew pathogen *Plasmopara viticola*. *Molecular Ecology*, 22, 2771-2786.
- Garbelotto M (2008) Molecular analysis to study invasions by forest pathogens: examples from Mediterranean ecosystems. *Phytopathologia Mediterranea*, 47, 183-203.
- Garbelotto M, Gonthier P (2013) Biology, epidemiology, and control of *Heterobasidion* species worldwide. *Annual Review of Phytopathology*, 51, in press.
- Garbelotto M, Linzer R, Nicolotti G, Gonthier P (2010) Comparing the influences of ecological and evolutionary factors on the successful invasion of a fungal forest pathogen. *Biological Invasions*, 12, 943-957.
- Gonthier P (2010) Invasion by pathogenic necrotrophic basidiomycota: shall we care? *Journal of Plant Pathology*, 92(4), S60.

- Gonthier P, Garbelotto M (2011) Amplified fragment length polymorphism and sequence analyses reveal massive gene introgression from the European fungal pathogen *Heterobasidion annosum* into its introduced congener *H. irregulare*. *Molecular Ecology*, 20, 2756-2770.
- Gonthier P, Warner R, Nicolotti G, Mazzaglia A, Garbelotto MM (2004) Pathogen introduction as a collateral effect of military activity. *Mycological Research*, 108, 468-470.
- Gonthier P, Garbelotto MM, Nicolotti G (2005) Seasonal patterns of spore deposition of *Heterobasidion* species in four forests of the western Alps. *Phytopathology*, 95, 759-767.
- Gonthier P, Nicolotti G, Linzer R, Guglielmo F, Garbelotto M (2007) Invasion of European pine stands by a North American forest pathogen and its hybridization with a native interfertile taxon. *Molecular Ecology*, 16, 1389-1400.
- Gonthier P, Anselmi N, Capretti P, Bussotti F, Feducci M, Garbelotto M, Giordano L, Guglielmo F, Honorati T, Lione G, Luchi N, Michelotti S, Michelozzi M, Nicolotti G, Paparatti B, Pollastrini M, Speranza S, Vettraino AM (2011) A comprehensive approach to study the impact and the epidemiology of the invasive forest pathogen *Heterobasidion irregulare* in Italy. *Journal of Plant Pathology*, 93(4), S34.
- Gonthier P, Lione G, Giordano L, Garbelotto M (2012) The American forest pathogen *Heterobasidion irregulare* colonizes unexpected habitats after its introduction in Italy. *Ecological Applications*, 22, 2135-2143.
- Gordon TR, Kirkpatrick SC, Petersen JC, Friel CJ (2006) Potential diversity in vegetative compatibility groupings in the California population of *Gibberella circinata*. *Mycological Research*, 110, 936-940.
- Grünwald NJ, Goodwin SB, Milgroom MG, Fry WE (2003) Analysis of genotypic diversity data for populations of microorganisms. *Phytopathology*, 93, 738-746.
- Hardy OJ, Vekemans X (2002) SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes*, 2, 618-620.
- Hedrick PW (1999) Perspective: highly variable loci and their interpretation in evolution and conservation. *Evolution*, 53, 313-318.
- Holdenrieder O, Pautasso M, Weisberg PJ, Lonsdale D (2004) Tree diseases and landscape processes: the challenge of landscape pathology. *Trends in Ecology and Evolution*, 19, 446-452.

- Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources*, 9, 1322-1332.
- Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, 23, 1801-1806.
- Johannesson HS, Stenlid J (2003) Molecular markers reveal genetic isolation and phylogeography of the S and F intersterility groups of the wood decay fungus *Heterobasidion annosum*. *Molecular Phylogenetics and Evolution*, 29, 94-101.
- Kofler R, Schlötterer C, Lelley T (2007) SciRoKo: a new tool for whole genome microsatellite search and investigation. *Bioinformatics*, 23, 1683-1685.
- Kretzer AM, Bruns TD (1999) Use of atp6 in fungal phylogenetics: an example from the Boletales. *Molecular Phylogenetics and Evolution*, 13, 483-492.
- Le Corre V, Kremer A (1998) Cumulative effects of founding events during colonisation on genetic diversity and differentiation in an island and stepping-stone model. *Journal of Evolutionary Biology*, 11, 495-512.
- Linzer RE, Otrosina WJ, Gonthier P, Bruhn J, Laflamme G, Bussieres G, Garbelotto M (2008) Inferences on the phylogeography of the fungal pathogen *Heterobasidion annosum*, including evidence of interspecific horizontal genetig transfer and of human-mediated, long-range dispersal. *Molecular Phylogenetics and Evolution*, 46, 844-862.
- Luikart G, Cornuet JM (1998) Empirical evaluation of a test for identifying recently bottlenecked populations from allele frequency data. *Conservation Biology*, 12, 228-237.
- Luikart G, Sherwin WB, Steele BM, Allendorf FW (1998) Usefulness of molecular markers for detecting population bottlenecks via monitoring genetic change. *Molecular Ecology*, 7, 963-974.
- Maloy OC (1997) White pine blister rust control in North America: a case history. *Annual Review of Phytopathology*, 35, 87-109.
- Milgroom MG, Wang K, Zhou Y, Lipari SE, Kaneko S (1996) Intercontinental population structure of the chestnut blight fungus, *Cryphonectria parasitica*. *Mycologia*, 88, 179-190.
- Muñoz-Díaz D, Rodrigo FS (2004) Spatio-temporal patterns of seasonal rainfall in Spain (1912-2000) using cluster and principal component analysis: comparison. *Annales Geophysicae*, 22, 1435-1448.

- Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York.
- Nei M, Tajima F (1981) DNA polymorphism detectable by restriction endonucleases. *Genetics*, 97, 145-163.
- Nei M, Maruyama T, Chakraborty R (1973) The bottleneck effect and genetic variability in populations. *Evolution*, 29, 1-10.
- Olson Å, Aerts A, Asiegbu F *et al.* (2012) Insight into trade-off between wood decay and parasitism from the genome of a fungal forest pathogen. *New Phytologist*, 194, 1001-1013.
- Otrosina WJ, Garbelotto M (2010) *Heterobasidion occidentale* sp. nov. and *Heterobasidion irregulare* nom. nov.: a disposition of North American *Heterobasidion* biological species. *Fungal Biology*, 114, 16-25.
- Pilcher HR (2004) Fungal stowaways. *Nature*, 428, 713.
- Piry S, Luikart G, Cornuet JM (1999) Bottleneck: a computer program for detecting recent reductions in the effective population size using allele frequency data. *Journal of Heredity*, 90, 502-503.
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, 155, 945-959.
- R Core Team (2012) *R: A Language and Environment for Statistical Computing*. R version 2.15.2. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, <http://www.R-project.org/>
- Roberts DW (2012) *Labdsv: Ordination and Multivariate Analysis for Ecology*. R package version 1.5-0, <http://CRAN.R-project.org/package=labdsv>
- Roccotelli A (2012) Characterization of wood decay fungi on citrus and forestry species. PhD Dissertation, Mediterranean University of Reggio Calabria.
- Rosenberg NA (2004) DISTRUCT: a program for the graphical display of population structure. *Molecular Ecology Notes*, 4, 137-138.
- Santini A, Ghelardini L, De Pace C *et al.* (2013) Biogeographical patterns and determinants of invasion by forest pathogens in Europe. *New Phytologist*, 197, 238-250.
- Shimizu M, Kosaka N, Shimada T, Nagahata T, Iwasaki H, Nagai H, Shiba T, Emi M (2002) Universal Fluorescent Labeling (UFL) method for automated microsatellite analysis. *DNA Research*, 9, 173-178.
- Simpson JA, May TW (2002) *Phaeolus schweinitzii* in Australia. *Australasian Plant Pathology*, 31, 99-100.

- Sokal RR, Oden NL (1978) Spatial autocorrelation in biology. I. Methodology. *Biological Journal of the Linnean Society*, 10, 199-228.
- Stenlid J (1985) Population structure of *Heterobasidion annosum* as determined by somatic incompatibility, sexual incompatibility, and isoenzyme patterns. *Canadian Journal of Botany*, 63, 2268-2273.
- Stenlid J, Karlsson J-O (1991) Partial intersterility in *Heterobasidion annosum*. *Mycological Research*, 95, 1153-1159.
- Vannini A, Franceschini S, Vettraino AM (2012) Manufactured wood trade to Europe: a potential uninspected carrier of alien fungi. *Biological Invasions*, 14, 1991-1997.
- Vekemans X, Hardy OJ (2004) New insights from fine-scale spatial genetic structure analyses in plant populations. *Molecular Ecology*, 13, 921-935.
- Vercken E, Fontaine MC, Gladieux P, Hood ME, Jonot O, Giraud T (2010) Glacial refugia in pathogens: European genetic structure of anther smut pathogens on *Silene latifolia* and *Silene dioica*. *PLoS Pathogens*, 6, e1001229.

Figure legends

Fig. 1. Map showing the spatial distribution of the six pine stands where isolates of *Heterobasidion irregulare* were sampled.

Fig. 2. Significant linear regressions (Pearson, $P < 0.05$) between the distance of sites from Castelfusano and their mean diversity indices (H and \hat{H}) calculated over seven variable loci.

Fig. 3. Allelic richness (a) and private allelic richness (b) of *Heterobasidion irregulare* as a function of standardized sample size for Anzio, Castelfusano, Circeo, Fregene-Coccia di Morto and Nettuno.

Fig. 4. Bar plots for $K = 2$, $K = 3$ and $K = 4$ showing the assignment values of genetic clusters of *Heterobasidion irregulare* populations in Italy (a), and scatterplot based on Principal Coordinates Analysis (PCoA) on isolates data (b). Isolates from different populations were marked with different symbols. Clusters identified by Hierarchical Cluster Analysis on principal coordinates are shown as groups of isolates encircled by different borders depending on the number of clusters.

Fig. 5. Spatial autocorrelation analysis of genetic and geographical distance in *Heterobasidion irregulare*. Moran's I -index averaged over SSR loci. Dashed lines: 95% CI; *** $P < 0.001$; * $P < 0.05$. Mean distances: 0.52, 2.42, 9.72, 29.26, 37.58, 44.53, 46.74, 58.58, 79.91, 91.70 km.

Supporting information 1. Isolates of *Heterobasidion irregulare* subjected to sequence analysis and related GenBank accession numbers of the sequenced loci.

Supporting information 2. Isolates of *Heterobasidion irregulare* subjected to SSR analysis, sample locations, and related number of repeats of the analyzed SSR loci.

Supporting information 3. Number of individuals analyzed per each genetic locus and site, and related different alleles detected.

Supporting information 4. Diversity indices calculated per each Italian and North American *Heterobasidion irregulare* population.

Supporting information 5. Magnitude of delta- K as a function of K (mean \pm sd over 10 replicates) for $K=2$ to 5, calculated with the formula $\text{delta-}K = \text{mean}[L''(K)] / \text{sd}[L(K)]$. A distinct peak at $K=3$ was indicative that the clustering level $K=3$ was optimal.

Supporting information 6. R script used to perform Principal Coordinates Analysis (PCoA), Hierarchical Cluster Analysis (HCA) and kernel density analysis. Instructions on how to run the script are reported in the script itself with classical notation (i.e. #instruction#). The content of files necessary to run the program is reported in paragraphs following the script.

Table 1. Loci sequenced in this study, related number of polymorphic sites and alleles detected among the individual analyzed.

Locus	Putative biological function	Amplicon length (bp)	N. of individuals	N. of polymorphic sites	N. of alleles	GB accession numbers [‡]
ACH*	Atrazine Chlorohydrolase/Guanine Deaminase-Amidohydrolase	696	34	12	6	JX982607
ATP†	Mitochondrial ATP synthase subunit 6	483	31	0	1	-
BTUB*	Beta-Tubulin	378	30	10	6	JX982608, JX982609, JX982610, JX982611, JX982612, JX982613
CAM*	Putative Calmodulin	136	31	2	3	-
EFA†	Elongation factor 1- α	335	53	7	5	JX982614
EPG*	Candidate Polygalacturonase Glycoside Hydrolase Family 28 protein	413	33	4	2	JX982615, JX982616
GPD†	Glyceraldehyde 3-phosphate dehydrogenase	535	31	5	6	JX982617
GST1*	Glutathione-S Transferase 1 gene	404	32	2	3	JX982618, JX982619, JX982620
ITS*	Internal Transcribed Spacer	571	32	0	1	JX982621
OMP*	Uridine 5'-Monophosphate Synthase/Orotate Phosphoribosyltransferase	642	29	0	1	-

* Primers for the amplification of these loci were reported in Gonthier & Garbelotto (2011).

† Primers for the amplification of these loci were reported in Kretzer & Bruns (1999), Johannesson & Stenlid (2003), and Gonthier *et al.* (2007).

‡ GeneBank accession numbers of novel alleles detected in this study.

Table 2. SSR loci analyzed, their location in the genome, primers used with related temperature of annealing, number of repeats and alleles among the individuals analyzed.

Locus	Genome location	Forward (F) and reverse (R) primers (5'-3')*	Ta (°C)†	N. of individuals	Amplicon length (bp)	Repeats	N. of repeats	N. of alleles
MS17a	Sc2: 564659	F: AGGAATCGTAGGGACGTTCG R: GTCGAACCCGAACCTGAACT	55	92	109-115	CTG	11-13	3
MS27a	Sc2: 3188791	F: CAGAAAGCGAAAACAGAGGAG R: GACGACGAACTCGTGGAAG	55	68	116	GAG	10	1
MS29a	Sc3: 153954	F: CACGTGACAGACACAGATGG R: CGCGTGATCCTCATTCCT	55	34	110-122	TCG	12-16	3
MS32a	Sc3: 935953	F: GAATGACCCACCCGACT R: CCCAAGAGAACCATCACCAT	55	64	111-123	ATG	8-12	5
MS37a	Sc3: 2232991	F: GAAGAGGAGAGAGGAGGAAAGAA R: ACACAGCGCCCCTTCACT	52	96	117-120	TCC	10-11	2
MS47a	Sc4: 1377840	F: CCTGTCCTTTTCCTCCGACT R: ACAGCTCAAGTCTGCCCTCT	55	78	115	CAG	10	1
Tri1	Sc1: 3573600	F: AGAACGAGAATGGGGAGGAT R: CGATCAAAAACGAGGAGGT	59	90	342-369	CTC	6-15	7
Tri4	Sc4: 1006374	F: AAGGGTGTCATTCGTTTCAGG R: CACCCTCCTCTTCATCCAA	59	95	124-157	GGT	2-13	7
Tri5	Sc5: 770074	F: TGAAGCTTTTCCAAGGCACT R: CTGCTGCTGCTGCTATTGTT	59	93	288-369	CAA	12-39	6
Tri11	Sc11: 868145	F: GCTCGTCCCTTCTCTCTCCT R: TTTGAACGAAATCCGAGGTC	58	88	387-399	TCA	5-9	4
Tri39	Sc39: 2355	F: GGGATCGAGTATGTGCGTC R: CCGGTCAACAGTCATTCCTT	59	96	280-283	TGT	10-11	2
Hexa1d	Sc1: 3300100	F: GACGTGCACCTTCATGCTC R: GGATTTGACGGCTCGAAGT	59	90	170-212	CACTCA	2-9	4

* A tag sequence (5'-GGTGCGACTCCTGGAG-3') is linked to all forward primers for universal labeling (Shimizu *et al.* 2002).

† Annealing temperature.

Table 3. Summary of the AMOVA results for the *Heterobasidion irregulare* populations in Italy.

Source of Variation	d.f.	Sum of squares	Variance components	% Variation
Among Populations	4	1205.601	9.256	6.30
Within Populations	90	12389.652	137.663	93.70
Total	94	13595.253	146.919	
Fixation Index Φ_{ST}	0.0630 ($P = 0.0429$)			

Table 4. Pairwise Φ_{ST} between populations of *Heterobasidion irregulare* in Italy. P -value calculated with 10000 permutations, and marked as significant with an asterisk after Bonferroni correction if $P < 0.0051$.

	Anzio	Castelfusano	Circeo	Fregene-Coccia di Morto	Nettuno
Anzio	-	0.0482	0.1099	0.0000	0.0000
Castelfusano	0.0482	-	0.2872*	0.0000	0.0000
Circeo	0.1099	0.2872*	-	0.1101	0.1050
Fregene-Coccia di Morto	0.0000	0.0000	0.1101	-	0.0000
Nettuno	0.0000	0.0000	0.1050	0.0000	-

Fig. 1

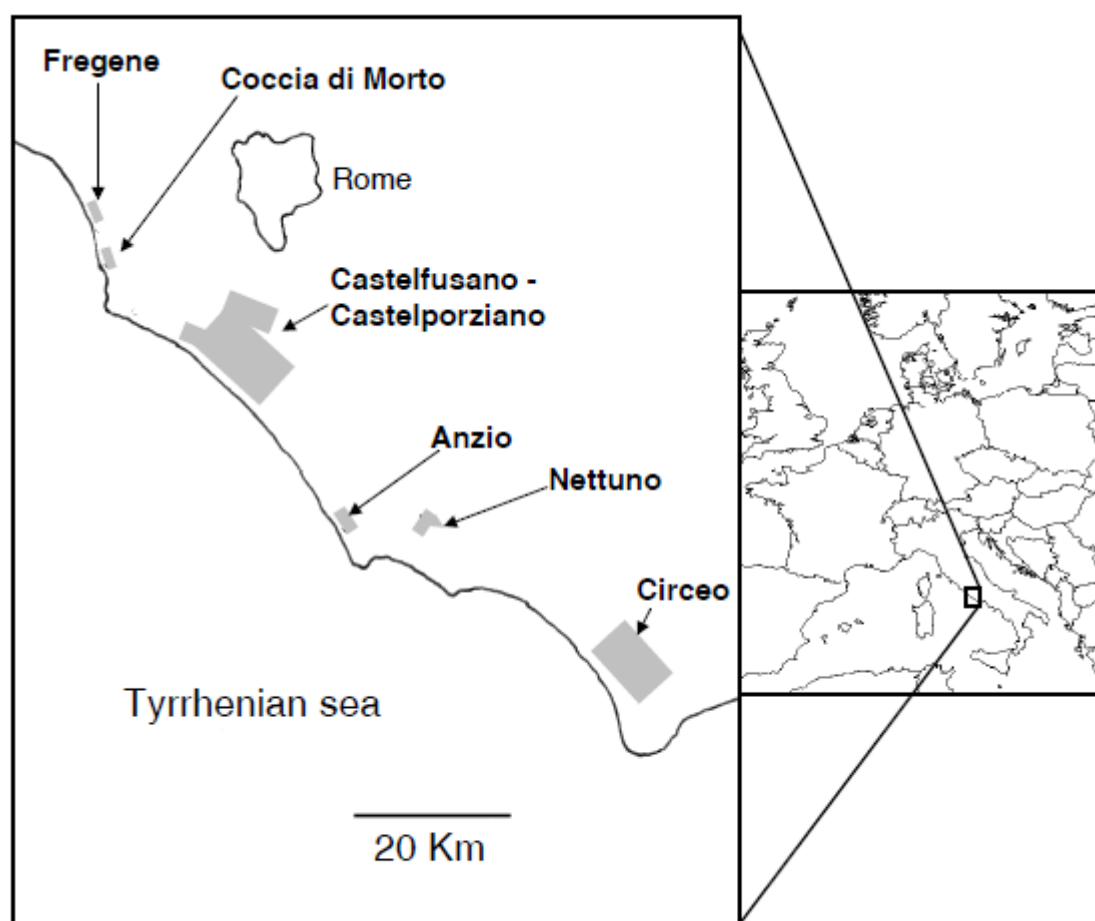


Fig. 2

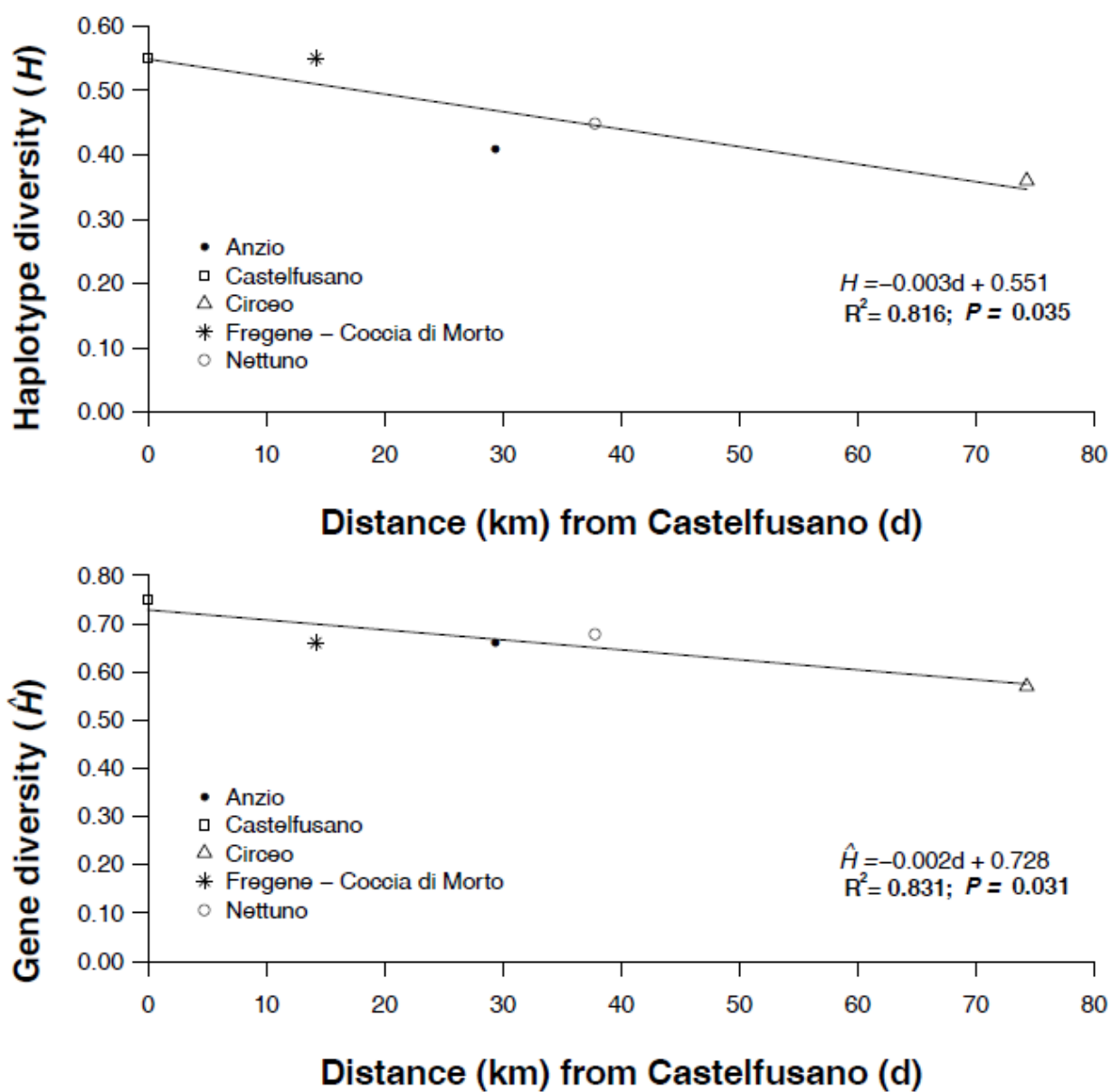


Fig. 3

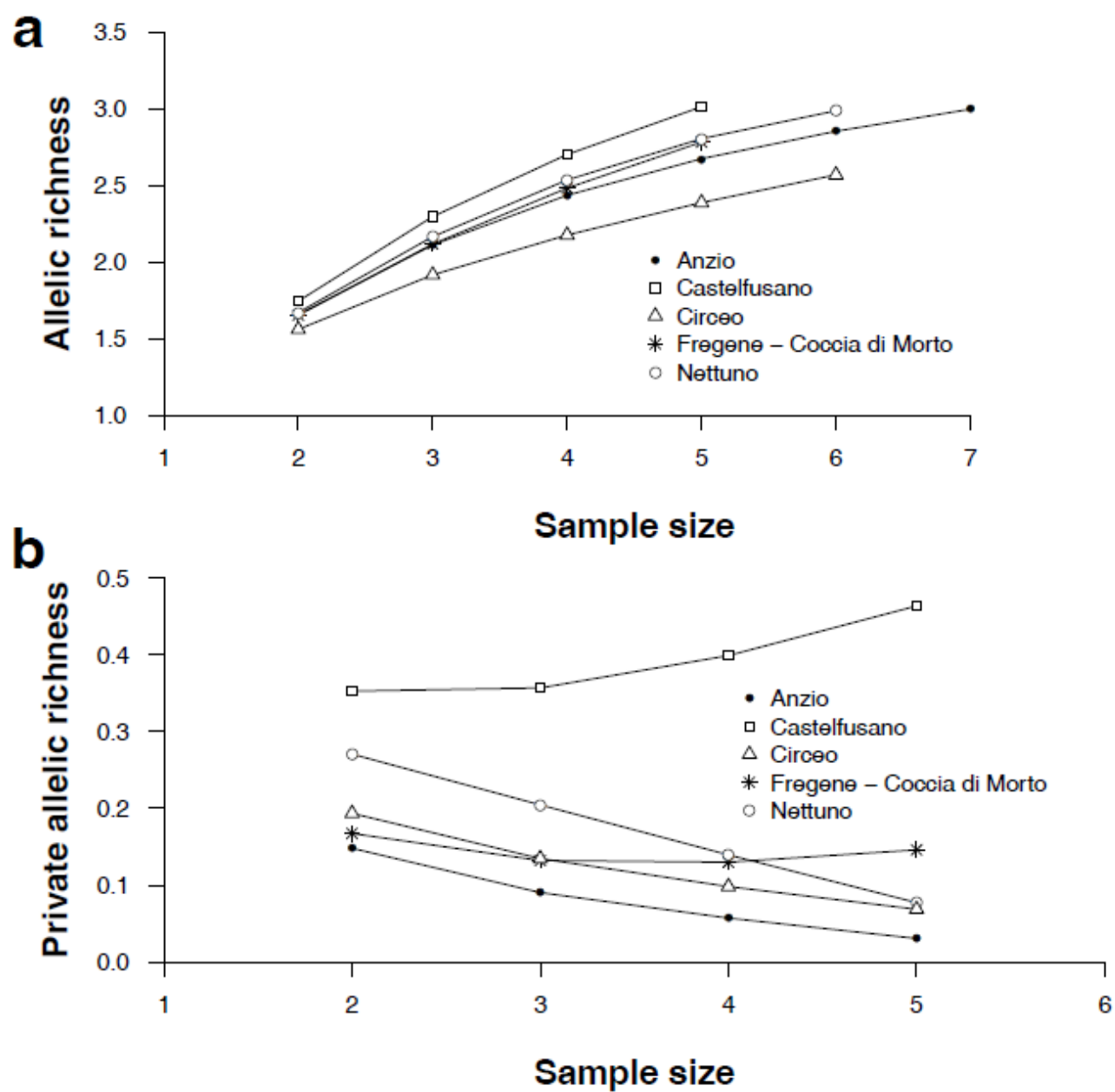


Fig. 4

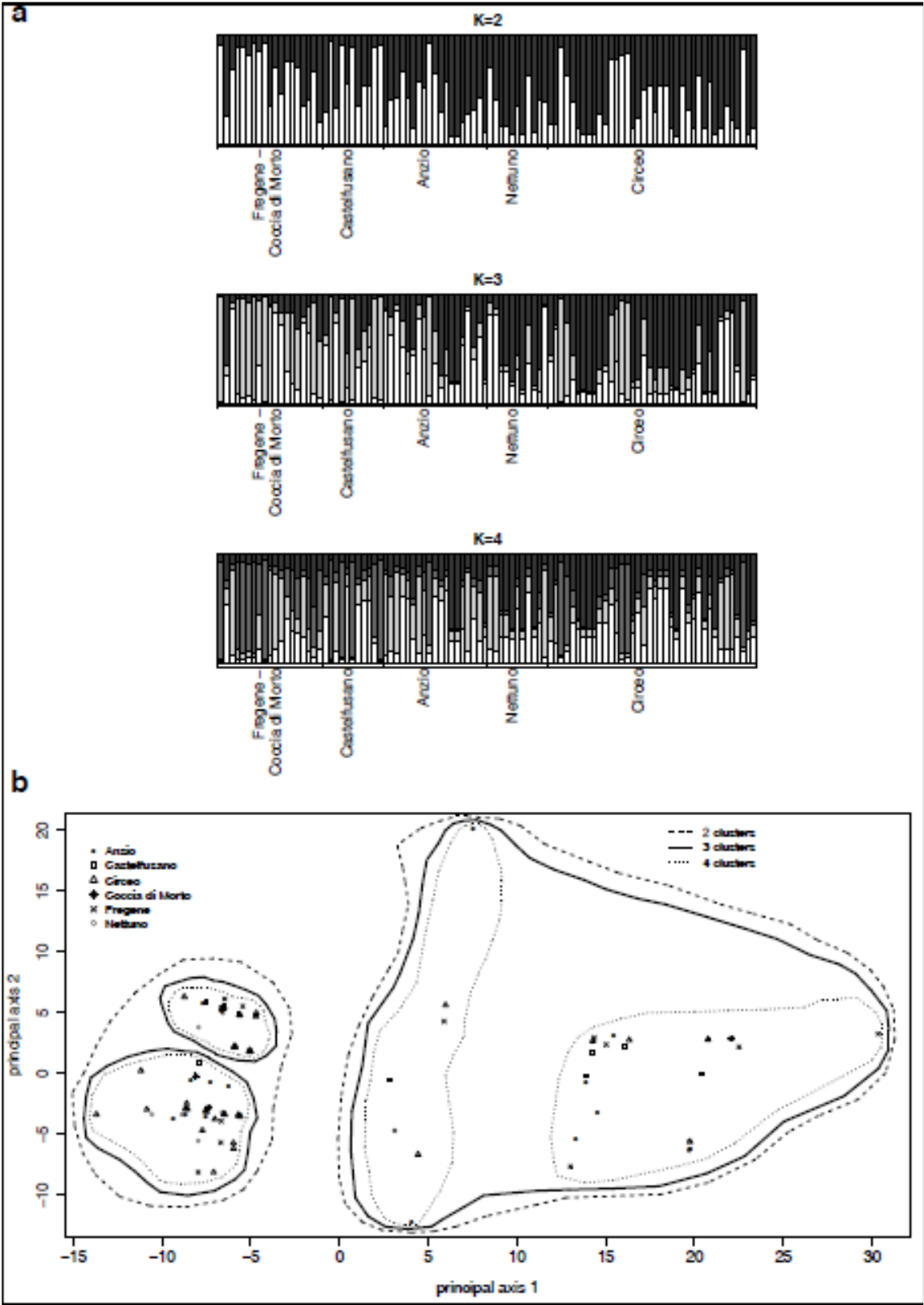


Fig. 5

